

Towards the definition of pathogenic microbe<sup>☆</sup>Yael Danin-Poleg<sup>a</sup>, Lilach Somer<sup>a</sup>, Lyora A. Cohen<sup>a</sup>, Eran Diamant<sup>a</sup>,  
Yniv Palti<sup>b</sup>, Yechezkel Kashi<sup>a,\*</sup><sup>a</sup> Department of Biotechnology and Food Engineering, Grand Water Research Institute, Technion, Israel Institute of Technology, Haifa 32000, Israel<sup>b</sup> National Center for Cool and Cold Water Aquaculture, USDA-ARS, 11861 Leetown Rd., Kearneysville, WV 25430, USA

Accepted 26 April 2006

## Abstract

Identification and typing of spoilage and pathogenic microorganisms have become major objectives over the past decade in microbiology. In food, strain typing is necessary to ensure food safety and for linking cases of foodborne infections to suspected items. Recent advances in molecular biology have resulted in the development of numerous DNA-based methods for discrimination among bacterial strains. Here, we present the use of Simple Sequence Repeats (SSR, or Microsatellites) for bacterial typing. SSRs are a class of short DNA sequence motifs that are tandemly repeated at a specific locus. Computer-based screen of the complete genomic DNA sequences of various prokaryotes showed the existence of tens of thousands well distributed SSR tracts. Mono Nucleotides Repeats (MNRs) are the majority of SSR tracts in bacteria, therefore selected MNR loci were analyzed for variation among strains belonging to three bacterial species: *Escherichia coli*, *Listeria monocytogenes* and *Vibrio cholerae*. High levels of polymorphism in the number of repeats was observed. The finding that most of the MNR tracts are variable in bacterial genomes, but stable at the strain level, allows the use of MNRs for bacterial strains identification. The variation in MNR tracts enables the separation between virulent and non-virulent strain groups and further discriminates among bacterial isolates, in the three tested bacterial species. The uncovered MNR polymorphism is important as a genome-wide source of variation, both in practical applications (e.g. rapid strain identification) and in evolutionary studies. This multi-locus MNR strategy could be applied for high throughput bacterial typing by assigning an “identity number” for each strain based on MNR variations. The developed typing technology should include the fingerprint database for large bacterial strain collections and a high throughput scanner. This accurate and rapid tool can have a major role in decreasing the incidences of food-related outbreaks and will contribute to limit epidemics.

© 2006 Elsevier B.V. All rights reserved.

**Keywords:** Bacterial typing; Simple sequence repeats; Microsatellite; Mononucleotide repeats; *Listeria monocytogenes*; *Escherichia coli*; *Vibrio cholerae*

## 1. Introduction

Foodborne pathogens are a major cause of human illness with estimated number of 76 million cases, 325,000 hospitalizations, and 5200 deaths in the United States each year and much higher proportion in the developed countries. In addition to the effect on human health, large percentages of the food products are lost by microbial spoilage. The modern food industry, based on international market, long shelf-life products and ready to eat food,

raises new microbial safety problems and demands. Consequently pathogenic microorganism detection and identification have become major objectives over the past decade in food and water microbiology. One specific problem is the discrimination among species and strains that may vary from normal pro-biotic micro-flora to virulent strains.

Bacterial typing allows distinguishing among strains within a species, to follow epidemics and route of contamination or to identify the source of spoilage. The ability to rapidly type an organism to the specific strain, particularly one that can pose a more serious threat, can lead to faster crisis response, limiting infection, aiding in treatment, and preventing potential spread of the organism. In addition we are witnessing the development of a new field of forensic microbiology. It requires accurate and reproducible technology that produces DNA typing signature or

<sup>☆</sup> This paper was presented at the 19th International ICFMH Symposium, Food Micro 2004, Potorož Slovenia, 12–16 September 2004.

\* Corresponding author. Tel.: +972 4 8293074; fax: +972 4 8293399.

E-mail address: [kashi@tx.technion.ac.il](mailto:kashi@tx.technion.ac.il) (Y. Kashi).

URL: <http://www2.technion.ac.il/biotech/> (Y. Kashi).

“identity number” allowing claiming in court that two isolates of bacteria are of the same clone (e.g. the search for the source of the recent Anthrax threat in the US (Read et al., 2002)). Such technology, like in human forensic typing, requires high discrimination power that facilitates the distinction among foodborne pathogens as well as among spoilage bacteria.

Classical bacterial isolation and identification is based on selective enrichment followed by plating on selective media. Species identification is mainly by biochemical characterization and strain identification is primarily based on serology. These methods are often labor-intensive and require days to complete and thus do not meet the requirement for rapid identification and typing. Recent advances in biotechnology have resulted in the development of numerous methods for typing of microorganisms such as ribotyping, PFGE and PCR-based methods e.g., repetitive extragenic palindrom (REP-PCR) (for review see: Olive and Bean, 1999; van Belkum, 2003; Wiedmann, 2002). Multi-Locus Sequence Typing (MLST) mainly of housekeeping genes is a rising, DNA-sequence based method, for bacterial strain typing (Clarke, 2002; Maiden et al., 1998). DNA sequencing provides far more variation per locus than any other method currently used for bacterial strain typing, and it provides a uniform platform for comparison between different laboratories and for database storage. Simple Sequence Repeats (SSRs), also termed VNTR (Variable Number of Tandem Repeats), are a class of short DNA sequence motifs that are tandemly repeated at a specific locus. A subgroup of SSRs are the mononucleotide repeats (MNR). A genome-wide screen of *Escherichia coli* demonstrated that polymorphism of MNRs is an important source of genome variation (Gur-Arie et al., 2000). The variation of SSR enabled their use for strain typing in several bacterial species (Keim et al., 1999, 2000; van Belkum et al., 1998). Recently, MLST of loci harboring mononucleotide repeats (MNRs) at non-coding regions were shown to contain much higher sequence variation (including Single Nucleotide Polymorphisms — SNPs) than housekeeping genes. This variation was shown to be more efficient and as reliable as MLST of housekeeping genes for strain discrimination and for inferring phylogenetic relationships in *E. coli* (Diamant et al., 2004).

Here we examined the suitability of the MNR–MLST approach for typing of three bacteria, one Gram-positive and two Gram-negative representatives, *L. monocytogenes* and *E. coli*, *Vibrio cholerae*. Those three are major foodborne pathogens and thus are important model for demonstrating pathogen identification methods. The three bacteria have wide spectrum of strains from harmless or pro-biotic bacteria to virulent, hence pointing to the importance of species (*L. monocytogenes*) and strain (*E. coli* and *V. cholerae*) typing. *E. coli* is a Gram-negative bacterium composed of numerous strains and serotypes (Ahmed et al., 1987; Ochman and Selander, 1984). The species includes commensal strains and a variety of pathogenic groups. EHEC serotype O157:H7 is one of the primary foodborne pathogenic threats in Europe and North America (Whittam et al., 1993). *Listeria* spp. are Gram-positive bacteria that include six classified species, among which only *L. monocytogenes* is pathogenic to human and responsible for listeriosis (Farber and Peterkin, 1991). *L. monocytogenes* is widely

distributed in the environment and common in food products. Moreover, the organism is able to survive and grow at most preservation conditions exclusively at refrigerator temperatures. Pregnant women, newborns, and immuno-compromised individuals are at particular risk for Listeriosis with high mortality rates of 20–30% even when appropriate medical care is provided (Vazquez-Boland et al., 2001). *V. cholerae* is the etiological agent of cholera. Its natural reservoir is the aquatic environment (Lipp et al., 2002; Wachsmuth et al., 1994). The pathogenic and non-pathogenic isolates of *V. cholerae* are divided to different serological groups: O1, O139, and the non-O1, non-O139, of which only O1 and O139 serogroups are associated with cholera pandemics.

The present study combines the variation of MNR and the MLST approach, termed MNR–MLST based method for accurate typing of three representative pathogens: *E. coli*, *L. monocytogenes* and *V. cholerae*.

## 2. Materials and methods

### 2.1. Bacterial strains

The study included representatives of three bacteria: *E. coli*, *Listeria* spp. and *V. cholerae*: (a) 27 pathogenic and non-pathogenic strains of *E. coli* (Table 2, Diamant et al., 2004), (b) 28 strains of *L. monocytogenes*, six strains of the remaining species of the *Listeria* genus; *L. innocua*, *L. seeligeri* (Fig. 3), *L. ivanovii* ATCC 19119, *L. welshimeri* ATCC 35897, *L. grayi* ATCC 19120, and *L. murrayi* ATCC 25401, and (c) 16 strains of *V. cholerae* (Fig. 1c), 9 clinical isolates belonging to O1 and O139 serogroups and 7 environmental, non-O1/non-O139, isolates.

### 2.2. Genome evaluation of SSR distribution

The complete genomic sequences of *E. coli* K12 (Blattner et al., 1997), *E. coli* O157:H7 (Hayashi et al., 2001), *L. monocytogenes* EGD-e, serotype 1/2a (Glaser et al., 2001) and *V. cholerae* El-Tor N16961 (Heidelberg et al., 2000) were obtained from the NCBI database (<http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi>). These chromosomes were screened for perfect mononucleotide repeats with number of repeats greater than five, using the “SSR” computer program (Gur-Arie et al., 2000). The program is available for downloading from the Technion’s Internet site at: <http://www.technion.ac.il/~anne/choice3.html>.

### 2.3. Loci and primer selection

Seven MNR loci were selected for the study, 5 *E. coli* MNR loci one *L. monocytogenes* locus and one *V. cholerae* locus. PCR primers were usually selected to amplify a specific intergenic locus, from one open reading frame (ORF) to the adjacent ORF as previously described (Gur-Arie et al., 2000). Loci were named after the downstream ORF. Primers were selected using the Gene Runner (Version 3.05).

*E. coli* primers were based on the K12 sequence (GenBank accession U00096). The primers used to amplify *serW* MNR locus were: Forward 5'-TTCACAGGTAACATACTCCAC-3',

and Reverse 5'-TTTGGTGAGGTGTCCGAG-3' giving a product of 193 bp of which 89 bp were sequenced. Primers for *b2345*, *ycgE*, *ycgW* and *yaiN* loci were previously published (Diamant et al., 2004).

*L. monocytogenes* primers were based on the published genome sequence (EGD-e, serotype 1/2a; GenBank accession NC\_003210). The primers used to amplify *Lmo0196* MNR locus

were: Forward 5'-GCTGCTAGTAAACAATGCT-3', and Reverse 5'-GGATAGAGAAGTATAAAAATG-3' giving a product of 435 bp of which 341 bp were sequenced.

*V. cholerae* primers were based on the published genome sequence of EL Tor N16961 (chromosome 1, GenBank accession NC\_002505). The primers used to amplify *VC1833-(T)*<sub>9</sub> MNR locus were: Forward 5'-ACACTTTATGAGCAAGGCTTC-3',

(a)		1								70
O42:H2	hap1	TCCAGGATAC	TCCCCCCCC	CCCC	TGGCTA	TTGTGCGCTC	ATACACTCAA	ATTAAAGATA	GGTTCTAAAT	
O26:H11	hap2	TCCAGGATAC	TCCCCCCCC	C---	TGGCTA	TTGTGCGCTC	ATACACTCAA	ATTAAAGATA	GGTTCTAAAT	
DH5a	hap3	TCCAGGATAC	TCCCCCCCC	---	TGGCTA	TTGTGCGCTC	ATACACTCAA	ATTAAAGATA	GGTTCTAAAT	
O157:H-	hap4	TCCAGGATAC	TCCCCCCCC	---	TGGCTA	TTGTGCGCTC	ATACACTCAA	ATTAAAGATA	GGTTCTAAAT	
O8:H9	hap5	TCCAGGATAC	TCACCCCCC	CC-	TGGCTA	TTGTGCGCTC	ATACACTCAA	ATTAAAGATA	GGTTCTAAAT	
BSR9b	hap6	TCCAGGATAC	TCACCCCC	---	TGGATA	TTGTGCGCTC	ATACACGCAA	ATTAAAGATA	GGTTCTAAAT	
O22:H8	hap7	TCCAGGATAC	TCACCCCC	---	TGGCTA	TTGTGCGCTC	ATACACTCAA	ATTAAAGATA	GGTTCTAAAT	
O26:H-	hap8	TCCAGGATAC	TCAACCCCC	---	TGGCTA	TTGTGCGCTT	ATACACTCAA	ATTAAAGATA	GGTTCTAAAT	
O157:H7	hap9	TCCAGGATAC	TCAACCCCC	---	TGGCTA	TTGTGCGCTC	ATACACGCAA	ATTAAAGATA	GGTTCTAAAT	
O157:H7 HER1057	hap9	TCCAGGATAC	TCAACCCCC	---	TGGCTA	TTGTGCGCTC	ATACACGCAA	ATTAAAGATA	GGTTCTAAAT	
O157:H7 HER1058	hap9	TCCAGGATAC	TCAACCCCC	---	TGGCTA	TTGTGCGCTC	ATACACGCAA	ATTAAAGATA	GGTTCTAAAT	
O157:H7 HER1261	hap9	TCCAGGATAC	TCAACCCCC	---	TGGCTA	TTGTGCGCTC	ATACACGCAA	ATTAAAGATA	GGTTCTAAAT	
O157:H7 HER1265	hap9	TCCAGGATAC	TCAACCCCC	---	TGGCTA	TTGTGCGCTC	ATACACGCAA	ATTAAAGATA	GGTTCTAAAT	
O157:H7 HER1266	hap9	TCCAGGATAC	TCAACCCCC	---	TGGCTA	TTGTGCGCTC	ATACACGCAA	ATTAAAGATA	GGTTCTAAAT	
O55:H7	hap9	TCCAGGATAC	TCAACCCCC	---	TGGCTA	TTGTGCGCTC	ATACACGCAA	ATTAAAGATA	GGTTCTAAAT	
O86:H18	hap10	TCCAGGATAC	TCAACCTCC	---	TGGCTA	TTGTGCGCTC	ATACACTCAA	ATTAAAGATA	GGCTCTAAAT	
Consensus		TCCAGGATAC	TC*CC*CC*	****	TGG*TA	TTGTGCGCT*	ATACAC*CAA	ATTAAAGATA	GG*TCTAAAT	

(b)		121	140	291					340
LM25H - 4a	hap1	AAAAATAGTT	TTTTTCGGA...	TATTAGTTTT	GTCA	TTTTTC	AGAAATAATT	GTATACAATT	ATTAGAGGGT
LM17 - 4c	hap1	AAAAATAGTT	TTTTTCGGA...	TATTAGTTTT	GTCA	TTTTTC	AGAAATAATT	GTATACAATT	ATTAGAGGGT
LM17/3- 1/2a	hap2	AAAAAGTTTT	TTTTTTGAA...	TATTTGATTT	ATCA	TTTTTC	GAAAGATAATT	GTATACAATA	TCTATAGGAT
LMW1 - 3a	hap2	AAAAAGTTTT	TTTTTTGAA...	TATTTGATTT	ATCA	TTTTTC	GAAAGATAATT	GTATACAATA	TCTATAGGAT
LMW16 - 3a	hap2	AAAAAGTTTT	TTTTTTGAA...	TATTTGATTT	ATCA	TTTTTC	GAAAGATAATT	GTATACAATA	TCTATAGGAT
LM14 - 1/2a	hap3	AAAAAGTTTT	TTTTTTGAA...	TATTTGATTT	ATCA	TTTTTC	GAAAGATAATT	GTATACAATA	TCTATAGGAT
LM16 - 1/2a	hap3	AAAAAGTTTT	TTTTTTGAA...	TATTTGATTT	ATCA	TTTTTC	GAAAGATAATT	GTATACAATA	TCTATAGGAT
LM15 - 1/2c	hap3	AAAAAGTTTT	TTTTTTGAA...	TATTTGATTT	ATCA	TTTTTC	GAAAGATAATT	GTATACAATA	TCTATAGGAT
LM24H - 1/2c	hap3	AAAAAGTTTT	TTTTTTGAA...	TATTTGATTT	ATCA	TTTTTC	GAAAGATAATT	GTATACAATA	TCTATAGGAT
LMW33 - 3c	hap3	AAAAAGTTTT	TTTTTTGAA...	TATTTGATTT	ATCA	TTTTTC	GAAAGATAATT	GTATACAATA	TCTATAGGAT
LMW52 - 3c	hap3	AAAAAGTTTT	TTTTTTGAA...	TATTTGATTT	ATCA	TTTTTC	GAAAGATAATT	GTATACAATA	TCTATAGGAT
LM8 - 4b	hap4	AAAAAGTTTT	TTTTTTGAA...	TATTTGATTT	ATCA	TTTTTC	GAAATAATT	GTATACAATA	TATATAGGGT
LM26H - 4b	hap4	AAAAAGTTTT	TTTTTTGAA...	TATTTGATTT	ATCA	TTTTTC	GAAATAATT	GTATACAATA	TATATAGGGT
LM31 - 1/2b	hap4	AAAAAGTTTT	TTTTTTGAA...	TATTTGATTT	ATCA	TTTTTC	GAAATAATT	GTATACAATA	TATATAGGGT
LM1 - 1/2b	hap4	AAAAAGTTTT	TTTTTTGAA...	TATTTGATTT	ATCA	TTTTTC	GAAATAATT	GTATACAATA	TATATAGGGT
LMW19 - 3b	hap4	AAAAAGTTTT	TTTTTTGAA...	TATTTGATTT	ATCA	TTTTTC	GAAATAATT	GTATACAATA	TATATAGGGT
LMW11 - 3b	hap4	AAAAAGTTTT	TTTTTTGAA...	TATTTGATTT	ATCA	TTTTTC	GAAATAATT	GTATACAATA	TATATAGGGT
L. innocua	hap5	AAAAATAGTT	TTTCTTTGAA...	TGTTAAAT	ATGATT	TTTCTC	AAAAATAATT	GTATACAATT	TACAGAGAAT
L. seeligeri	hap5	AAAAATAGTT	TTTCTTTGAA...	TGTTAAAT	ATGATT	TTTCTC	AAAAATAATT	GTATACAATT	TACAGAGAAT
Consensus		AAAAA*TTT	TTT*T**GAA...	T*TT***TT	*T*A	TTT**C	**A*ATAATT	GTATACAAT*	***A*AG**T

(c)		ctxA haplotype	121								190
Serotype											
O1 Inaba-1 <sup>a</sup>	+ hap1	TAGACATGCC	CGTTTTTTTT	T-CGATGGGC	ATGTCTATTT	TCAGCACAAC	AGACAGCTTA	CTTACGCGTT			
O1 Ogawa 185 <sup>a</sup>	+ hap1	TAGACATGCC	CGTTTTTTTT	T-CGATGGGC	ATGTCTATTT	TCAGCACAAC	AGACAGCTTA	CTTACGCGTT			
O1 Inaba 391-93 <sup>b</sup>	+ hap1	TAGACATGCC	CGTTTTTTTT	T-CGATGGGC	ATGTCTATTT	TCAGCACAAC	AGACAGCTTA	CTTACGCGTT			
O1 Inaba 390-93 <sup>b</sup>	- hap1	TAGACATGCC	CGTTTTTTTT	T-CGATGGGC	ATGTCTATTT	TCAGCACAAC	AGACAGCTTA	CTTACGCGTT			
O1 Ogawa ElTor4 <sup>a</sup>	+ hap1	TAGACATGCC	CGTTTTTTTT	T-CGATGGGC	ATGTCTATTT	TCAGCACAAC	AGACAGCTTA	CTTACGCGTT			
O1 Ogawa ElTor20 <sup>a</sup>	- hap1	TAGACATGCC	CGTTTTTTTT	T-CGATGGGC	ATGTCTATTT	TCAGCACAAC	AGACAGCTTA	CTTACGCGTT			
O1 Ogawa 41 <sup>a</sup>	+ hap1	TAGACATGCC	CGTTTTTTTT	T-CGATGGGC	ATGTCTATTT	TCAGCACAAC	AGACAGCTTA	CTTACGCGTT			
O1 Inaba 155314 <sup>a</sup>	+ hap1	TAGACATGCC	CGTTTTTTTT	T-CGATGGGC	ATGTCTATTT	TCAGCACAAC	AGACAGCTTA	CTTACGCGTT			
O139 58-99 <sup>b</sup>	+ hap1	TAGACATGCC	CGTTTTTTTT	T-CGATGGGC	ATGTCTATTT	TCAGCACAAC	AGACAGCTTA	CTTACGCGTT			
O9 Shafdan <sup>c</sup>	- hap2	TAGACATGCC	CGTTTTTTT	-CGATGGGC	ATGTCTATTT	TCAGCACAAC	AGACAGCTTA	CTTACGCGTT			
O2 Kishon183 <sup>c</sup>	- hap3	TAGACATGCC	CATTTTTTAA	GGCGATGGGC	ATGTCTATTT	TCAGCACAAC	AGACAGCTTA	CTTACGCGTT			
O10 Zanzibar C22 <sup>c</sup>	- hap3	TAGACATGCC	CATTTTTTAA	GGCGATGGGC	ATGTCTATTT	TCAGCACAAC	AGACAGCTTA	CTTACGCGTT			
O79 Tivon 145 <sup>c</sup>	- hap3	TAGACATGCC	CATTTTTTAA	GGCGATGGGC	ATGTCTATTT	TCAGCACAAC	AGACAGCTTA	CTTACGCGTT			
O141 475-00 <sup>b</sup>	+ hap3	TAGACATGCC	CATTTTTTAA	GGGGATGGGC	ATGTCTATTT	TCAGCACAAC	AGACAGCTTA	CTTACGCGTT			
O22 Japan <sup>a</sup>	- hap4	TAGACATGCC	CGTTTTTTT	-CGATGGGC	ATGTCTATTT	TCAGCACAAC	AGACAGCTTA	CTTACGCGTT			
O141 849-95 <sup>b</sup>	- hap5	TAGACATGCC	CATTTTAAAT	GGTGATGGGC	ATGTCTATTT	TCAGCACAAC	AGACAGCTTA	CTTACGCGTT			
Consensus		TAGACATGCC	C*TTTT**	***GATGGGC	ATGTCTATTT	TCAGCACAAC	AGACAGCTTA	CTTACGC*T			

Fig. 1. Sequence alignments at three MNR loci of *E. coli* (a), *Listeria* spp. (b) and *Vibrio cholerae* (c). Analyzed loci are *B2345*, *Lmo0196* and *VC1833-(T)*<sub>9</sub>, respectively. [(a) DH5a presents the same haplotype as the published sequence of *E. coli* K12. (b) LM14 is *L. monocytogenes* (LM) EGD-e 1/2a. No product was amplified in *L. ivanovii*, *L. welshimeri*, *L. grayi*, and *L. murrayi*. (c) <sup>a</sup>Collection of the Israeli Ministry of Health. All O1 isolates are clinical. <sup>b</sup>Arakawa E. — Department of Bacteriology, National Institute of Infectious Diseases, Tokyo, Japan. <sup>c</sup>Environmental, isolated from Chironomides egg mass and flying adults near water sources or waste stabilization pond (Broza and Halpern, 2001; Halpern et al., 2004). Source of isolation is indicated.]

and Reverse 5'-ATCACTTATCCCGCTTAATCG-3' giving a product of 263 bp of which 193 bp were sequenced.

#### 2.4. DNA extraction and PCR amplification

Pure cultures high quality genomic DNA extraction was used through the study as previously described (Diamant et al., 2004; Gur-Arie et al., 2000).

The PCR reaction mixture contained: 0.2 mM dNTPs; 10  $\mu$ M each of the forward and reverse primers; 0.25 u Taq polymerase (Super Nova — JMR holding, Kent, England); 1 $\times$  buffer (1.5 mM MgCl<sub>2</sub>); 50 ng template DNA in a total volume of 25  $\mu$ l. The reaction was carried out in a PCR thermocycler (HYBAID Omn-E, Ashford, UK), as follows: 95 °C for 5 min; five cycles of 45 s at 95 °C, 45 s at  $T_m$  and 45 s at 72 °C; 20 cycles of 45 s at 95 °C, 45 s at  $T_m$  – 5 °C, and 45 s at 72 °C; and a final step of 72 °C for 7 min. The annealing temperatures ( $T_m$ ) were 58 °C, 54 °C, 60 °C for *serW*, *Lmo0196* and *VC1833-(T)<sub>9</sub>*, respectively and for *b2345*, *ykgE*, *ycgW* and *yaiN* loci as previously published (Diamant et al., 2004). PCR amplification products were analyzed by agarose gel (2%) electrophoresis, and observed by UV fluorescence.

#### 2.5. DNA sequencing

PCR products were purified using a QIAquick PCR purification kit (QIAGEN). 50 ng of the purified product were sequenced on both strands using the BigDye™ terminator Cycle Sequencing kit (Perkin-Elmer Applied Biosystems) and loaded onto an ABI 310 automated sequencer. Only sequences with complete agreement between the two strands were used for further analysis. Multiple alignments of the sequences were performed with the Sequence Navigator™ program (version 1.0.1, ABI).

#### 2.6. Data and statistical analysis

Two strategies were used to assess the variation at MNR loci: sequence comparison and haplotype analysis.

##### 2.6.1. Haplotype analysis

Haplotype (microhaplotype) analysis is an a-parametric analysis in which each haplotype gets different allele number at a locus. Haplotypes include all sequence variations: MNR and SNP polymorphisms (MNR–MLST, Fig. 1). An additional allele was counted where there was no amplification product (e.g. Table 2; Haplotype no. 1).

Diversity index was calculated as  $1 - \sum P_{ij}^2$ , where  $P_{ij}$  is the frequency of the  $j$ th allele at the  $i$ th locus. The data for all genotypes were scored as present (1) or absent (0) for each of the haplotype at specific locus.

Genetic relations among strains were inferred from the haplotype data of each of the three pathogens. The data were used to calculate the Nei coefficient of association (Nei, 1973) and generate the corresponding matrix using the SAS 8.02. This matrix was used to create dendrograms based on UPGMA (the unweighted pair-group method with arithmetic average) method using MEGA 2.1 software (Kumar et al., 2001).

##### 2.6.2. Sequence comparison

The phylogenetic analyses for *Listeria* spp. and *V. cholerae* were inferred from the sequences data of *Lmo0196* locus and *VC1833-(T)<sub>9</sub>* locus, respectively. Multiple alignments of the sequences were performed with the Sequence Navigator™ (version 1.0.1, ABI) or by using Clustal-X (Jeanmougin et al., 1998). The alignments files, in PIR format were inserted to the Gapcoder software (Young and Healy, 2003) for the indel variations analysis. The output files (NEXUS format) including the indel variations coded as presence/absence were transformed to FASTA format via SEAVIEW software (Galtier et al., 1996) following transformation of the presence/absence to A/C respectively (0=A, 1=C). These files were converted to MEGA format and used to evaluate genetic relationships construct dendrograms by the UPGMA method (MEGA 2.1, Kumar et al., 2001). Gaps were treated as missing data, pairwise. Bootstrap confidence values were based on 1000 simulated dendrograms.

### 3. Results and discussion

#### 3.1. SSR distribution

Genome wide screen of *E. coli* (K12 and O157:H7), *L. monocytogenes* (EGD-e, serotype 1/2a) and *V. cholerae* (El-Tor N16961) revealed tens of thousands of perfect SSR tracts. The current scan was for SSR with repeats number larger than three, and MNR larger than five bp. The majority of the SSR (*E. coli* — 88%, *L. monocytogenes* — 94% and *V. cholerae* — 90%) were found to be mononucleotide repeats (MNRs). Hence we concentrate on their analysis. MNR are evenly distributed and highly abundant (Table 1) throughout the bacterial genomes with average appearance of one MNR tract every 241 bp, 225 bp, 110 bp, and 224 bp, in *E. coli* K12, *E. coli* O157:H7, *L. monocytogenes* and *V. cholerae* genomes, respectively. Although the genome of *L. monocytogenes* is the smallest genome analyzed in the current study, MNR frequency was found to be the highest. Length distribution of MNR in the genome is similar in the three bacterial species tested despite of the differences in genome size. Most of the MNR tracts in the four bacterial genomes are short and did not exceed 9 bp (Table 1), as previously reported for other bacterial genomes (Field and Wills, 1998; Gur-Arie et al., 2000; Metzgar et al., 2001). Thus, short MNR are highly distributed and the most abundant SSR in bacterial genome as previously reported (Gur-Arie et al., 2000; Klevytska et al., 2001). MNR were previously found to be polymorphic among *E. coli* strains despite their short length (Diamant et al., 2004; Gur-Arie et al., 2000; Metzgar et al., 2001). Five MNR loci were chosen for polymorphism study in *E. coli* (Table 2), one locus in *Listeria* and one in *V. cholerae*.

#### 3.2. Sequence variation at MNR loci

High polymorphism was found among the strains of the three bacterial species in the tested MNR loci (Table 2, Fig. 1). The current study included different strains of *E. coli*, *L. monocytogenes*, *Listeria* spp. and *V. cholerae* from diverse

Table 1  
Number of mononucleotide repeat (MNR) loci along the genome of *E. coli*, *L. monocytogenes* and *V. cholerae*

Bacterial species	Genome size Mbp	Number of tandem MNR repeats (bp)						
		5	6	7	8	9	10	11–13
<i>E. coli</i> K12 MG1655	4.64	13880	4119	1003	215	19	1	0
<i>E. coli</i> O157:H7	5.50	17394	5349	1351	308	29	0	1 <sup>a</sup>
<i>L. monocytogenes</i> EGD-e serotype 1/2a	2.94	17641	6565	2331	243	16	1	1 <sup>b</sup>
<i>V. cholerae</i> O1 biovar eltor str. N16961 Chromosomes 1+2	4.03	13014	3925	862	143	21	2	2 <sup>b</sup>

<sup>a</sup> Number of repeats=13.

<sup>b</sup> Number of repeats=11.

origin. Sequence analysis of the MNR loci showed that variations were mostly due to variability in repeat number of the MNR tract in comparison with the flanking sequences, as demonstrated in Fig. 1. Additional SNPs were observed in the flanking sequences. These multiple sequence variations formed between four to ten haplotypes (micro-haplotypes) per locus, most of which are MNR alleles. For example, 10 haplotypes were observed among *E. coli* strain at *b2345* locus, all of them harbor different MNR alleles (Fig. 1a). Four of the five haplotypes identified among *L. monocytogenes* strain and four of the eight haplotypes (5 presented) identified among *V. cholerae* strains contained various MNR alleles (Fig. 1b and c, respectively). Likewise, six haplotypes were observed when all 34 strains of *Listeria* spp. were tested, 5 of them were MNR alleles. The high variation of MNR loci was found in rather short DNA fragments (ranging from 200 to 350 base pairs) making these loci mostly informative for multi-locus sequence typing (MLST) in comparison with the relatively long DNA stretches that are needed to obtain such variation via the MLST method of housekeeping genes (Enright and Spratt, 1999).

The original MLST method is based on sequence variation at housekeeping genes that are highly conserved, therefore usually present in all strains but have limited variation (Farfan et al., 2002; Noller et al., 2003; Salcedo et al., 2003). Sequence analysis methods such as MLST require data for all tested strains in a specific locus. The current haplotype analysis method considers additional variations at the priming site that result in a non-amplified (“Null”) allele. In such cases, where no product was amplified at the tested locus, additional haplotype allele variation was counted. The haplotype analysis provides 100% type-ability of any strain at any selected locus, by combining both sequence variations and product amplification results, in comparison to the commonly used sequence based methods that consider only sequence variations (Maiden et al., 1998; Urwin and Maiden, 2003). For example, only this analysis facilitated the use of the variation at *serW* locus as a base for typing since 12 of the *E. coli* strains did not amplify any product at this locus (Table 2). Likewise in *Listeria* MNR-sequence variation among strains and species is applicable at the *Lmo0196* locus, only upon including an un-amplified product in the analysis. It enabled the analysis of

all 34 *Listeria* strains yielding seven haplotypes with diversity index of DI=0.77. The variations in the tested MNR loci yielded high diversity values, ranging from 0.66 to 0.85, among strains of the three bacterial species (e.g., *E. coli*, Table 2) indicate their potential use for bacterial typing.

Comparison between bacterial species shows that in *L. monocytogenes* sequence variations of the MNR tracts are more likely to be the outcome of single base substitutions, while in *V. cholerae* and *E. coli* the variation is mostly due to insertion/deletions of the repeated nucleotide (Fig. 1), leading to size variation of the DNA fragment. Furthermore, in *Listeria* many additional SNPs are observed at the flanking sequences compared to few in *V. cholerae* and *E. coli* (data not shown). Analysis of the entire sequence variations of these MNR loci (MNR–MLST) should enable detailed discrimination between isolates.

### 3.3. Phylogentic analysis and discrimination based on MNR variation

The sequence variations (MNR and SNP polymorphisms) at the MNR loci were used for discrimination and phylogenetic analyses among strains of the three bacteria. Two strategies

Table 2  
Haplotypes of 27 *E. coli* strains at 5 MNR loci

Strain <sup>a</sup>	Locus				
	<i>ykgE</i>	<i>yaiN</i>	<i>b2345</i>	<i>ycgW</i>	<i>serW</i>
SR9b (B)	4	3	6	7	2
SR9c (B)	4	3	6	7	2
O78:H <sup>−</sup> (ETEC)	5	3	2	6	1
O111:H <sup>−</sup> (EHEC)	5	3	5	7	1
O113:H2 (EHEC)	5	2	4	7	1
O157:H <sup>−</sup> (EHEC)	5	3	4	4	2
DH5a (K12)	1	5	2	5	2
W4100 (K12)	1	5	2	5	2
O55:H7 (EPEC)	6	1	8	2	3
O111ac:H <sup>−</sup> (EPEC)	5	5	2	5	1
W3110 (K12)	5	5	2	5	2
O22:H8 (EHEC)	5	3	5	9	1
O26:H11 (EHEC)	5	3	3	8	1
O42:H2 (EHEC)	5	2	3	7	1
O157:H7 (EHEC)	6	1	8	1	3
O157:H7-1057 (EHEC)	6	1	8	1	3
O157:H7-1059 (EHEC)	6	1	8	1	3
O157:H7-1261 (EHEC)	6	1	8	1	3
O157:H7-1265 (EHEC)	6	1	8	1	3
O157:H7-1266 (EHEC)	6	1	8	1	3
86:H10 (ETEC)	7	1	9	7	4
O86:H18 (ETEC)	2	6	10	10	1
O8:H9 (ETEC)	5	2	7	7	1
O9:H33 (ETEC)	5	5	1	3	2
O153:H <sup>−</sup> (ETEC)	5	4	4	6	1
O26:H <sup>−</sup> (EPEC)	3	3	11	7	1
O127:H21 (EPEC)	5	2	2	9	1
Total haplotypes number	7	6	11	10	4
Diversity index (DI)	0.69	0.77	0.85	0.82	0.67

Haplotype number 1 at all loci presents absent of the expected amplification PCR product.

<sup>a</sup> For further information concerning source of strains and sequence data for loci: *b2345*, *ykgE*, *ycgW* and *yaiN* see Diamant et al. (2004).

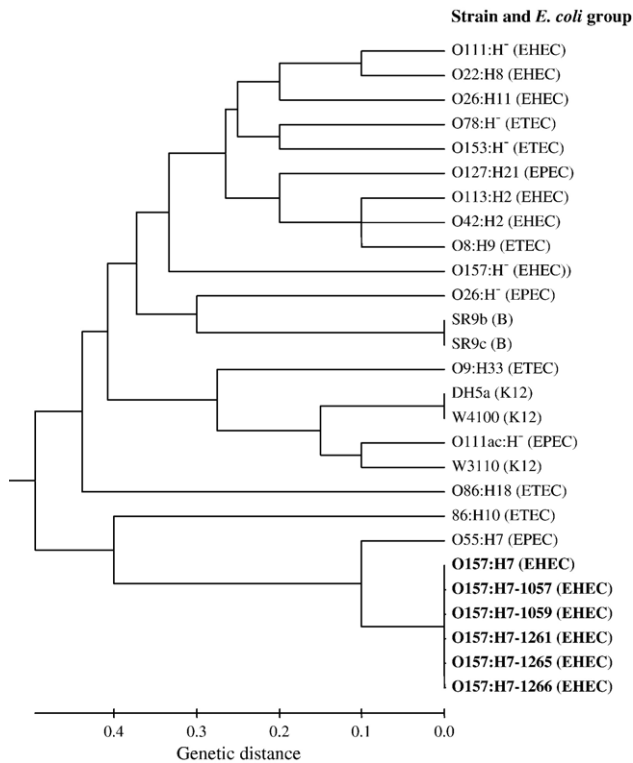


Fig. 2. Phylogenetic tree of 27 *E. coli* strains constructed by UPGMA cluster analysis of haplotypes data at five MNR loci (Table 2; 38 polymorphic points).

were employed for the analysis; sequence comparison and haplotype analyses. A clear discrimination among strains of the three bacterial species was demonstrated by both strategies.

Genetic relations among the 27 *E. coli* strains were inferred from the haplotype data at five MNR loci, based on 38 polymorphic points (loci × number of alleles in each locus). Genetic distances among *E. coli* isolates ranged from 0.00 for the very close isolates to 1.00 for the most distant isolates with a mean genetic distance of  $0.79 \pm 0.28$ . Cluster analysis presented in Fig. 2, enable the discrimination among 20 of the 27 *E. coli* strains. O157:H7 serological group, the major food threat, was clearly separated from the other strains. O157:H7 isolates were clustered together with the O55:H7 isolate, supporting the findings that O55:H7 and O157:H7 have recently evolved from a common ancestor (Whittam, 1996). The six O157:H7 isolates exhibited the same pattern indicating close genetic relations and low diversity, in agreement with other methods (Kudva et al., 2002; Noller et al., 2003).

The human pathogen *L. monocytogenes* was clearly separated from the other *Listeria* species. Fig. 3 presents cluster analysis inferred from sequence comparison at *Lmo0196* MNR locus. A clear discrimination among *L. monocytogenes* strains was demonstrated. *L. monocytogenes* was separated into the three known Lineages consisting of serotype groups: 1/2b, 3b and 4b — Lineage I; 1/2a, 1/2c, 3a and 3c — Lineage II; 4a and 4c — Lineage III (Fig. 3). Lineage I includes the most virulent strains that are the casual agents of most foodborne outbreaks of which, serotype 4b is responsible for 70% of the Listeriosis patients. The current separation is highly significant (70% of Bootstrap values >90, Fig. 3) even-though it is based on one locus of 341 bp. Similar results were obtained using sequence information from single genomic loci in *sigB* gene (Moorhead et al., 2003) and *inlB* gene (Ericsson et al., 2000), both are contingency genes with high polymorphic level (Moxon et al.,

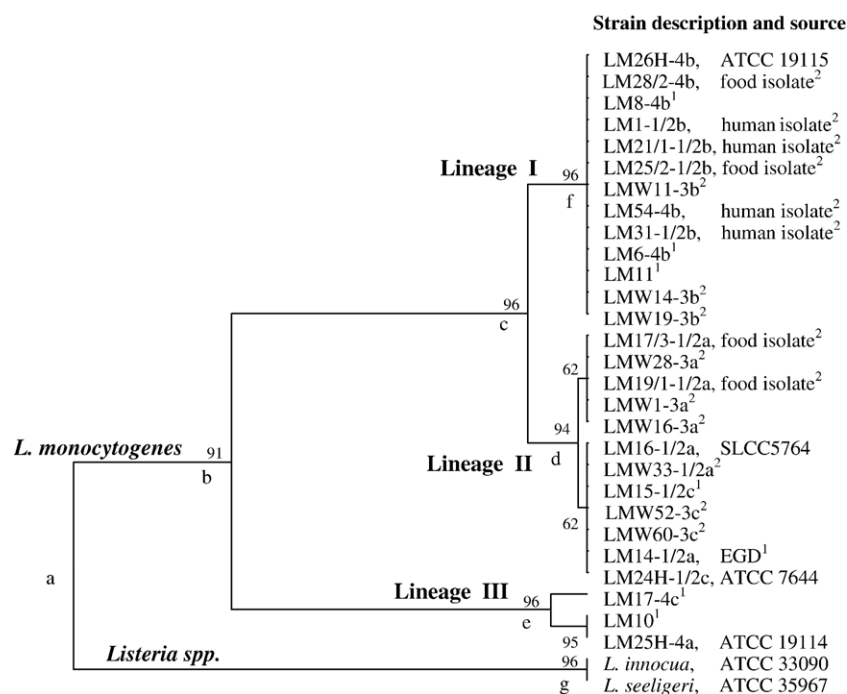


Fig. 3. Phylogenetic tree of 28 *L. monocytogenes*, *L. innocua* and *L. seeligeri* strains based on the DNA sequence at the *lmo0196* MNR locus. The tree was constructed with the UPGMA method. The numbers at the nodes are bootstrap confidence values based on 1000 replicates. <sup>1</sup>Food Microbiology laboratory, Wayne State University, Detroit, Michigan, USA <sup>2</sup>Ministry of Health, Central Laboratories, Jerusalem, Israel.

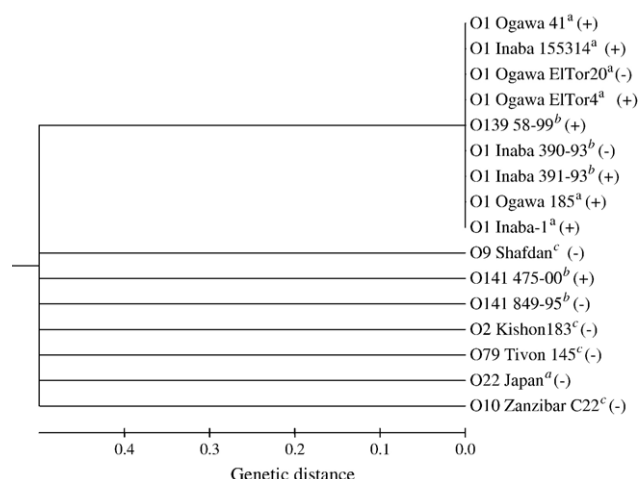


Fig. 4. Phylogenetic tree of 16 *V. cholerae* strains constructed by UPGMA cluster analysis of haplotypes data of the *VC1833-(T)<sub>9</sub>* MNR locus. +/– stands for presence/absence of *ctxA* gene.

1994). These results indicate that analysis of additional MNR loci would enable higher discrimination between closely related strains, like in other typing methods based on multi-locus information e.g., MLST and PFGE (Brosch et al., 1994; Salcedo et al., 2003).

The clinical *V. cholerae* strains were clearly separated from the environmental isolates based on haplotype analysis of sequence information at the *VC1833-(T)<sub>9</sub>* MNR locus (Fig. 4). Furthermore, we could discriminate among all seven environmental, non-O1/non-O139, isolates indicating that they belong to different origins (Singh et al., 2001). All six O1 isolates (*ctxA*-positive and negative) and O139 isolate *ctxA*-positive, were grouped together indicating, like other studies (Singh et al., 2001) that O1 and O139 strains have the same clonal origin. Analysis of 193 bp at the *VC1833-(T)<sub>9</sub>* MNR locus showed that all clinical had the same haplotype (Fig. 1c). Further discrimination among the clinical strain should be achieved by analysis of additional loci.

High discrimination among strains of both Gram-negative and Gram-positive bacteria was demonstrated using sequence base typing of MNR loci. The combination of MNR and single nucleotide variations at these loci enabled strain discrimination by using a few loci. Additional loci would enlarge the discrimination ability between closely related strains as demonstrated in *E. coli*. Furthermore, the discrimination ability of the MNR–MLST method is high relative to the short length of the determined sequence, ranging from 89–341 bp for a locus, in comparison to the long sequence stretches needed for typing with house keeping genes (Enright and Spratt, 1999). Thus the MNR–MLST method provides much higher variation per base. In addition, the presented haplotype analysis that combines both sequence variations and product amplification results including un-amplified product, provides 100% type-ability of any strain at any selected locus.

Although MNR are hyper-polymorphic sequences they were shown here to be stable at the strain level. The high homology among strains in the specific groups such as, *E. coli* O157:H7

and *V. cholerae* O1, is an additional indication for MNR low mutation rate. Such homology among the pathogenic strains can be applied for discrimination between virulent and non-virulent bacteria. Additional MNR site would contribute to the discrimination power, which is a function of the number of the analyzed loci.

### 3.4. Conclusion

The multi-locus MNR strategy could be applied for rapid bacterial typing by assigning an “identity number” for each strain based on sequence variation at MNR loci. Short DNA sequences of high variation content such as MNR loci can be analyzed by high-throughput methods like allele specific hybridization, Pyrosequencing and mass spectrometry. Thus, MNR–MLST method allows testing large number of samples in short time, supplying precise answers, in real time, of the specific contaminating bacteria. Such identification will have a major role in decreasing the incidences of food-related outbreaks and will contribute to limit epidemiology and the number of recalls. MNR–MLST provides a platform for comparison between laboratories as the sequence data can be easily transferred and stored, and it is suitable for bacterial identification and typing since it includes the advantages of specificity, sensitivity, rapid, low cost, and could be automated enabling large routine sampling.

### Acknowledgement

The authors thank Drs. Thomas and Beth Whittam, Michigan State University; Leora Shelef, Wayne State University, USA; Dr Arakawa E, National Institute of Infectious Diseases, Tokyo, Japan; the Ministry of Health Central Laboratories Jerusalem and Haifa Department, Israel, for providing the samples of the bacterial strains and serotyping. This research was supported by the Grand Water Research Institute, Technion; The Israeli Water Commission and by Israeli Science Foundation grant 1005697.

### References

- Ahmed, R., Bopp, C., Borczyk, A., Kasatiya, S., 1987. Phage-typing scheme for *Escherichia coli* O157:H7. J. Infect. Dis. 155, 806–809.
- Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., Shao, Y., 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277, 1453–1474.
- Brosch, R., Chen, J., Luchansky, J.B., 1994. Pulsed-field fingerprinting of listeriae: identification of genomic divisions for *Listeria monocytogenes* and their correlation with serovar. Appl. Environ. Microbiol. 60, 2584–2592.
- Broza, M., Halpern, M., 2001. Pathogen reservoirs. Chironomid egg masses and *Vibrio cholerae*. Nature 412, 40.
- Clarke, S.C., 2002. Nucleotide sequence-based typing of bacteria and the impact of automation. BioEssays 24, 858–862.
- Diamant, E., Palti, Y., Gur-Arie, R., Cohen, H., Hallerman, E.M., Kashi, Y., 2004. Phylogeny and strain typing of *Escherichia coli*, inferred from variation at mononucleotide repeat loci. Appl. Environ. Microbiol. 70, 2464–2473.
- Enright, M.C., Spratt, B.G., 1999. Multilocus sequence typing. Trends Microbiol. 7, 482–487.

- Ericsson, H., Unnerstad, H., Mattsson, J.G., Danielsson-Tham, M.L., Tham, W., 2000. Molecular grouping of *Listeria monocytogenes* based on the sequence of the *inlB* gene. *J. Med. Microbiol.* 49, 73–80.
- Farber, J.M., Peterkin, P.I., 1991. *Listeria monocytogenes*, a food-borne pathogen. *Microbiol. Rev.* 55, 476–511.
- Farfan, M., Minana-Galbis, D., Fuste, M.C., Loren, J.G., 2002. Allelic diversity and population structure in *Vibrio cholerae* O139 Bengal based on nucleotide sequence analysis. *J. Bacteriol.* 184, 1304–1313.
- Field, D., Wills, C., 1998. Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. *Proc. Natl. Acad. Sci. U. S. A.* 95, 1647–1652.
- Galtier, N., Gouy, M., Gautier, C., 1996. SEAVIEW and PHYLO\_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.* 12, 543–548.
- Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A., Baquero, F., Berche, P., Bloecker, H., Brandt, P., Chakraborty, T., Charbit, A., Chetouani, F., Couve, E., de Daruvar, A., Dehoux, P., Domann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L., Dussurget, O., Entian, K.D., Fsihi, H., Portillo, F.G., Garrido, P., Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J., Jackson, D., Jones, L.M., Kaerst, U., Kreft, J., Kuhn, M., Kunst, F., Kurapkat, G., Madueno, E., Maitournam, A., Vicente, J.M., Ng, E., Nedjari, H., Nordsiek, G., Novella, S., de Pablos, B., Perez-Diaz, J.C., Purcell, R., Rammel, B., Rose, M., Schlueter, T., Simoes, N., Tierrez, A., Vazquez-Boland, J.A., Voss, H., Wehland, J., Cossart, P., 2001. Comparative genomics of *Listeria* species. *Science* 294, 849–852.
- Gur-Arie, R., Cohen, C.J., Eitan, Y., Shelef, L., Hallerman, E.M., Kashi, Y., 2000. Simple sequence repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism. *Genome Res.* 10, 62–71.
- Halpern, M., Broza, Y.B., Mittler, S., Arakawa, E., Broza, M., 2004. Chironomid egg masses as a natural reservoir of *Vibrio cholerae* non-O1 and non-O139 in freshwater habitats. *Microb. Ecol.* 47, 341–349.
- Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M., Shinagawa, H., 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res.* 8, 11–22.
- Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleishmann, R.D., Nierman, W.C., White, O., 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 406, 477–483.
- Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G., Gibson, T.J., 1998. Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* 23, 403–405.
- Keim, P., Klevytska, A.M., Price, L.B., Schupp, J.M., Zinser, G., Smith, K.L., Hugh-Jones, M.E., Okinaka, R., Hill, K.K., Jackson, P.J., 1999. Molecular diversity in *Bacillus anthracis*. *J. Appl. Microbiol.* 87, 215–217.
- Keim, P., Price, L.B., Klevytska, A.M., Smith, K.L., Schupp, J.M., Okinaka, R., Jackson, P.J., Hugh-Jones, M.E., 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J. Bacteriol.* 182, 2928–2936.
- Klevytska, A.M., Price, L.B., Schupp, J.M., Worsham, P.L., Wong, J., Keim, P., 2001. Identification and characterization of variable-number tandem repeats in the *Yersinia pestis* genome. *J. Clin. Microbiol.* 39, 3179–3185.
- Kudva, I.T., Evans, P.S., Perna, N.T., Barrett, T.J., Ausubel, F.M., Blattner, F.R., Calderwood, S.B., 2002. Strains of *Escherichia coli* O157:H7 differ primarily by insertions or deletions, not single-nucleotide polymorphisms. *J. Bacteriol.* 184, 1873–1879.
- Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17, 1244–1245.
- Lipp, E.K., Huq, A., Colwell, R.R., 2002. Effects of global climate on infectious disease: the cholera model. *Clin. Microbiol. Rev.* 15, 757–770.
- Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., Spratt, B.G., 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U. S. A.* 95, 3140–3145.
- Metzgar, D., Thomas, E., Davis, C., Field, D., Wills, C., 2001. The microsatellites of *Escherichia coli*: rapidly evolving repetitive DNAs in a non-pathogenic prokaryote. *Mol. Microbiol.* 39, 183–190.
- Moorhead, S.M., Dykes, G.A., Cursons, R.T., 2003. An SNP-based PCR assay to differentiate between *Listeria monocytogenes* lineages derived from phylogenetic analysis of the *sigB* gene. *J. Microbiol. Methods* 55, 425–432.
- Moxon, E.R., Rainey, P.B., Nowak, M.A., Lenski, R.E., 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* 4, 24–33.
- Nei, M., 1973. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. U. S. A.* 70, 3321–3323.
- Noller, A.C., McEllistrem, M.C., Stine, O.C., Morris, J.G.J., Boxrud, D.J., Dixon, B., Harrison, L.H., 2003. Multilocus sequence typing reveals a lack of diversity among *Escherichia coli* O157:H7 isolates that are distinct by pulsed-field gel electrophoresis. *J. Clin. Microbiol.* 41, 675–679.
- Ochman, H., Selander, R.K., 1984. Standard reference strains of *Escherichia coli* from natural populations. *J. Bacteriol.* 157, 690–693.
- Olive, D.M., Bean, P., 1999. Principles and applications of methods for DNA-based typing of microbial organisms. *J. Clin. Microbiol.* 37, 1661–1669.
- Read, T.D., Salzberg, S.L., Pop, M., Shumway, M., Umayam, L., Jiang, L., Holtzapple, E., Busch, J.D., Smith, K.L., Schupp, J.M., Solomon, D., Keim, P., Fraser, C.M., 2002. Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science* 296, 2028–2033.
- Salcedo, C., Arreaza, L., Alcalá, B., De La Fuente, L., Vazquez, J.A., 2003. Development of a multilocus sequence typing method for analysis of *Listeria monocytogenes* clones. *J. Clin. Microbiol.* 41, 757–762.
- Singh, D.V., Matte, M.H., Matte, G.R., Jiang, S., Sabeena, F., Shukla, B.N., Sanyal, S.C., Huq, A., Colwell, R.R., 2001. Molecular analysis of *Vibrio cholerae* O1, O139, non-O1, and non-O139 strains: clonal relationships between clinical and environmental isolates. *Appl. Environ. Microbiol.* 67, 910–921.
- Urwin, R., Maiden, M.C., 2003. Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol.* 11, 479–487.
- van Belkum, A., 2003. High-throughput epidemiologic typing in clinical microbiology. *Clin. Microbiol. Infect.* 9, 86–100.
- van Belkum, A., Scherer, S., van Alphen, L., Verbrugh, H., 1998. Short-sequence DNA repeats in prokaryotic genomes. *Microbiol. Mol. Biol. Rev.* 62, 275–293.
- Vazquez-Boland, J.A., Kuhn, M., Berche, P., Chakraborty, T., Dominguez-Bernal, G., Goebel, W., Gonzalez-Zorn, B., Wehland, J., Kreft, J., 2001. *Listeria* pathogenesis and molecular virulence determinants. *Clin. Microbiol. Rev.* 14, 584–640.
- Wachsmuth, K., Olsvik, O., Evins, G.M., Popovic, T., 1994. *Vibrio cholerae* and Cholera: Molecular to Global Perspective. ASM Press, Washington DC.
- Whittam, 1996. Genetic variation and evolutionary processes in natural population of *Escherichia coli*, In: Neidhardt, F.C. (Ed.), *Escherichia coli* and *Salmonella* Cellular and Molecular Biology, Second edn. ASM Press, Washington D.C., pp. 2708–2720.
- Whittam, T.S., Wolfe, M.L., Wachsmuth, I.K., Orskov, F., Orskov, I., Wilson, R.A., 1993. Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. *Infect. Immun.* 61, 1619–1629.
- Wiedmann, M., 2002. Subtyping of bacterial foodborne pathogens. *Nutr. Rev.* 60, 201–208.
- Young, N.D., Healy, J., 2003. GapCoder automates the use of indel characters in phylogenetic analysis. *BMC Bioinformatics* 4, 6.